REMARKS

Claims 11, 12, 16, 22-27, 32-34, 44, 50-54, 56, 58-69, 73-81, and 83-109 are pending and under examination. Claims 11, 16, 22-27, 32, 34, 44, 50-53, 56, 58-68, 74-79, 81, 83, 84 and 86-99 have been amended. Support for the amendments can be found throughout the specification and claims as originally filed. In particular, support for the amendment to claims 11, 16, 25-27, 32, 34, 44, 50-53, 56, 58-68, 74-79, 81, 83, 84 and 86-99 can be found in the specification on page 14, lines 20-25 and page 35, line 31 to page 32, line 6. Claims 22-24 have been amended to depend on claim 16 and support for the amendment can be found in the originally filed claim 21. Accordingly, these amendments do not raise an issue of new matter and entry thereof is respectfully requested.

Rejection Under 35 U.S.C. §102

The rejection of claims 11, 16, 24-27, 32, 34, 44, 50-54, 56, 58-61, 67, 68, 75-79, 81, 83-86, 88, 89, 91, 92, 94, 95, 97, 98 and 100-109 under 35 U.S.C. §102(b) as allegedly being anticipated by Turner et al., Breast Cancer Res. Treatment 46:69 (1997) as evidenced by Krajewski et al., Endocrine-Related Cancer 6:29-40 (1999), and the Breastcancer.org website entitled "Stages of Breast Cancer" (www.breastcancer.org/symptoms/diagnosis/staging.jsp) submitted as Exhibit 2 in the response filed July 23, 2008, is respectfully traversed. Applicant respectfully maintains, for the reasons of record, that Turner et al. does not anticipate the claimed methods and further provides the following remarks.

The Office Action states on page 4 that the IC studied by Turner et al. is early stage, as evidenced by the disclosure of Krajewski et al. The Office Action then states that early stage invasive cancer of breast includes stages I and II breast cancer, as evidenced by Exhibit 2. Furthermore, the Office Actions states that stage I breast cancer is defined as invasive breast cancer (cancer cells are breaking through to or invading neighboring normal tissue) in which the tumor measures up to two centimeters and no lymph nodes are involved, and stage II breast cancer is defined as invasive breast cancer in which the tumor measures at least two centimeters but not more than five centimeters, or cancer has spread to the lymph nodes under the arm on the same side as the breast cancer. Thus, Turner et al. allegedly studied BAG-1 expression in breast cancer which is infiltrating [invasive] but has spread no further than the lymph nodes local to

breast. Applicant respectfully submits that 1) the Office is improperly using the disclosure of Krajewski et al. and Exhibit 2 to support the anticipation rejection, and 2) Turner et al. do not teach, expressly or inherently, the claimed methods relating to stage I or stage II of breast cancer or the prognostic methods as claimed.

Applicant respectfully submits that the Office is improperly using the disclosure of Krajewski et al. and Exhibit 2 to support the anticipation rejection over the disclosure of Turner et al. A rejection under 35 U.S.C. 102 over multiple references has been held to be proper when the extra references are cited to: (A) prove the primary reference contains an "enabled disclosure," (B) explain the meaning of a term used in the primary reference; or (C) show that a characteristic not disclosed in the reference is inherent (see MPEP 2131.01). It appears that the Office is using the disclosure of Krajewski et al. and Exhibit 2 to explain the meaning of the term "invasive carcinoma," also referred to as "IC," as disclosed by Turner et al. Extrinsic evidence may be used to explain but not expand the meaning of terms and phrases used in the reference relied upon as anticipatory of the claimed subject matter (*In re Baxter Travenol Labs*., 952 F.2d 388, 21 USPQ2d 1281 (Fed. Cir. 1991), emphasis added). See also *Scripps Clinic & Research Foundation V. Genetech, Inc.*, 927 F.2d 1565, 18 USPQ2d 1001 (Fed. Cir. 1991), wherein the court concludes, while it is sometimes appropriate to consider extrinsic evidence to explain the disclosure of a reference, the role of extrinsic evidence is to educate the decision-maker to what the reference means to persons of ordinary skill in the field of the invention, not to fill gaps in the reference. Applicant respectfully submits that the Office is using the disclosure of Krajewski et al. in combination with Exhibit 2 to fill in gaps and expand the disclosure of Turner et al.

At best, Turner et al. discloses the 10-year overall survival (OS) and distant disease free survival (DDFS) for patients with overexpression of cytoplasmic BAG-1 in invasive carcinoma (IC) specimens was 75% and 70%, respectively, as compared with 62% and 35% for tumors with low cytoplasmic BAG-1 levels ($p = 0.06$). With respect to invasive carcinoma as described by Turner et al., Applicant refers to the evidence of record, specifically Exhibit 1 (a printout from the Breastcancer.org website entitled "Non-Invasive or Invasive Breast Cancer?" (www.breastcancer.org/symptoms/diagnosis/invasive.jsp)), Exhibit 2, *supra*, and Exhibit 3 (Markman, Basic Cancer Medicine pgs. 35-37 (1997)) submitted with the response filed July 23, 2008. Exhibit 1 defines invasive (or infiltrating) breast cancers to be cancers that have started to

break through normal breast tissue barriers and invade surrounding area. Exhibit 2 explicitly discloses that breast cancer at stage I, stage II, stage III or stage IV is considered to be invasive. Additionally, according to the tumor-node-metastasis (TNM) staging system described in Exhibit 3, there are thirteen distinguishable TNM stage groupings within stage I to stage IV of breast cancer (see page 36 of Exhibit 3). Thus, Applicant submits that one skilled in the art would understand that invasive carcinoma, as taught by Turner, could include stage I, II, III, or IV breast cancers or any of the thirteen TNM stage groupings. Applicant also submits that, based on the description in Turner et al. and what was well known in the art, one skilled in the art would not know which stage of breast cancer the IC samples of Turner et al. belong and certainly would not know if the samples were stage I or stage II.

Applicant further submits that Turner et al. do not disclose or use the term "early stage." The Office argues that Krajewski et al. discloses on page 36, first column, lines 4-9; "Using a monoclonal antibody that we generated against BAG-1 protein and which recognizes all three of the known BAG-1 isoforms (Takayama et al. 1998), we examined the expression and intracellular location of BAG-1 in early-stage breast cancers (Turner et al. 1997)." Thus, the Office concludes that the IC studied by Turner is allegedly early stage. Applicant respectfully submits that the Office is using the disclosure of Krajewski et al. to fill in gaps and expand the disclosure of Turner et al., which is an improper use of extrinsic evidence to support an anticipation rejection. Absent the disclosure of Krajewski et al., Turner et al. are silent as to whether the breast cancer samples were stage I, II, III or IV or any of the thirteen TNM stage groupings. This is in contrast to the claimed methods, wherein the individual has stage I breast cancer in which the cancer is infiltrating but has no lymph node involvement or the individual has stage II breast cancer in which the cancer is infiltrating but has spread no further than the lymph nodes local to breast. Additionally, the Office's own admission on page 10, lines 16-17 of the Office Action issued November 24, 2009 states "Turner et al. do not expressly disclose that the invasive carcinoma of breast cancer include state I and/or stage II cancer." (emphasis added). Thus, Turner et al. do not teach, expressly or inherently, the claimed methods relating to stage I or stage II of breast cancer.

Furthermore, the method described by Turner et al. is a retrospective study, whereas the claimed methods are prognostic. A retrospective study, as described by Turner et al., looks

backwards and examines samples obtained from patients with known outcomes that is established at the start of the study. The only prognostic suggestion offered by Turner et al. relates to the subcellular localization of BAG-1, which may have prognostic importance with respect to survival of breast cancer patients (see last sentence of Abstract). This is in contrast to the claimed prognostic methods, which are directed towards methods for prognosis of disease-free or overall survival, methods for predicting the risk of tumor recurrence or spread, methods for screening a patient to determine the risk of tumor metastasis or chance of survival, or methods for determining the proper course of treatment for a patient, wherein the individual has stage I breast cancer in which the cancer is infiltrating but has no lymph node involvement or the individual has stage II breast cancer in which the cancer is infiltrating but has spread no further than the lymph nodes local to breast. Thus, the methods disclosed by Turner et al. are retrospective, whereas the claimed methods are prognostic with respect to stage I or stage II breast cancer.

According to MPEP § 2131, a claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference. As discussed above, the Office is improperly using the disclosure of Krajewski et al. and Exhibit 2 to support the anticipation rejection. Furthermore, Turner et al. do not teach, expressly or inherently, the claimed methods relating to stage I or stage II of breast cancer or the prognostic methods as claimed. Absent such teachings, Applicant respectfully submits that Turner et al. cannot anticipate the claims. Accordingly, Applicant respectfully requests that this rejection be withdrawn.

Rejections Under 35 U.S.C. §103

Maintained Rejections

The rejection of claims 11, 12, 16, 24-27, 32-34, 44, 50-54, 56, 58-61, 67-69, 75-81, 83-86, 88, 89, 91, 92, 94, 95, 97, 98 and 100-109 under 35 U.S.C. §103(a) as allegedly obvious over Turner et al., *supra* in view of Sano et al. (US Patent 5,665,539) as evidenced by Krajewski et al., *supra*, and Exhibit 2, *supra*, is respectfully traversed. Applicant respectfully maintains, for the reasons of record, that the claimed methods are unobvious over Turner et al., alone or in combination with Sano et al and further provides the following remarks.

The Office Action states on page 6 that Applicant's previous arguments are not persuasive for the reasons set forth in the 102(b) rejection. Applicant respectfully submits that the disclosure of Turner et al. alone or in combination with Sano et al. as evidenced by Krajewski et al. and Exhibit 2, do not teach or suggest the claimed prognostic methods relating to stage I or stage II of breast cancer and the level of cytosolic BAG-1 protein expression.

As discussed in the remarks above, the Office Action states on page 4 that the IC studied by Turner et al. is early stage, as evidenced by the disclosure of Krajewski et al. The Office Action then states that early stage invasive cancer of breast includes stages I and II breast cancer, as evidenced by Exhibit 2. Furthermore, the Office Actions states that stage I breast cancer is defined as invasive breast cancer (cancer cells are breaking through to or invading neighboring normal tissue) in which the tumor measures up to two centimeters and no lymph nodes are involved, and stage II breast cancer is defined as invasive breast cancer in which the tumor measures at least two centimeters but not more than five centimeters, or cancer has spread to the lymph nodes under the arm on the same side as the breast cancer. Thus, Turner et al. allegedly studied BAG-1 expression in breast cancer which is infiltrating [invasive] but has spread no further than the lymph nodes local to breast. Applicant respectfully disagrees for the following reasons.

At best, Turner et al. disclose, using a retrospective study, that the 10-year overall survival (OS) and distant disease free survival (DDFS) for patients with overexpression of cytoplasmic BAG-1 in invasive carcinoma (IC) specimens was 75% and 70%, respectively, as compared with 62% and 35% for tumors with low cytoplasmic BAG-1 levels ($p = 0.06$). Applicant maintains that a skilled artisan working in the cancer field would have understood that this relatively small differential in overall survival rate between specimen over- or under-expressing BAG-1 with p=0.06 is not significant enough to conclusively suggest that BAG-1 can be a reliable diagnostic tool for predicting survival, let alone being a diagnostic tool for predicting survival of patients with stage I or stage II of breast cancer as claimed. Applicant further provides herewith Exhibit A, Zar, Biostatistical Analysis, Prentice-Hall, Inc., Chapter 6, pgs. 79-86, (1999). Zar discloses on pages 81-82:

> As explained below, a probability of 5% [$p = 0.05$] or less is commonly used as the criterion for rejection of H_0 [the null

> hypothesis]. The probability used as the criterion for rejection is called the *significance level*,* denoted by α (the lowercase Greek letter, alpha). [*see page 81, last paragraph, emphasis added*]
>
> By experience, and hence by convention, an α of 0.05 is usually considered to be a "small enough" chance of committing a Type I error [a rejection of a null hypothesis when it is in fact true], while not being so small as to result in "too large a chance" of a Type II error [not rejecting the null hypothesis when it is in fact false]. But there is nothing sacrosanct about the 0.05 level. Although it is the most widely used significance level, researchers may decide for themselves whether it is more important to minimize one type of error or the other [*see page 82, last paragraph, emphasis added*]

In other words, by Turner et al. identifying a *p* value of 0.06, one skilled in the art would most likely conclude that there was no statistically significant correlation between OS and DDFS for invasive carcinoma patients with overexpression of cytoplasmic BAG-1. The evidence provided by Exhibit A corroborates Dr. Reed's statement that the patient survival difference of cytosolic staining of BAG-1 cited in Turner et al. for invasive carcinoma (IC) was not statistically significant (p = 0.06) (see Reed Declaration submitted previously as Exhibit 2 with the response filed December 13, 2006). Accordingly, Turner et al. describe a retrospective study that showed no statistically significant correlation between invasive carcinoma patients survival and overexpression of cytoplasmic BAG-1. It is the Applicant, in the present invention, that discovered 1) a statistically significant correlation between 10-year OS and DMFS (distant metastasis-free survival) for patients with overexpression of BAG-1 protein in stages I and II of breast cancer that was 90% and 84%, respectively, as compared with 40% and 40% for those with low BAG-1 levels (p<0.001) (see page 35 of the application as filed, lines 5-14, and Figure 1) and 2) patients whose tumors contained high levels of cytosolic BAG-1 protein are more likely to enjoy long-term survival and freedom from tumor recurrence or spread and distant metastases, compared to those with tumors containing low levels of cytosolic BAG-1 (see page 35 of the application as filed, line 31 to page 36, line 3).

Furthermore, the disclosure of Turner et al. as evidenced by Krajewski et al., actually teaches away from the claimed methods. At best, Krajewski et al. disclose higher levels of BAG-1 nuclear immunostaining (>20%) correlated with longer OS among patients with early stage breast cancer (p<0.001) (see page 36, first column, lines 19-23). This is in contrast to the

claimed prognostic methods, which are directed towards methods for prognosis of disease-free or overall survival, methods for predicting the risk of tumor recurrence or spread, methods for screening a patient to determine the risk of tumor metastasis or chance of survival, or methods for determining the proper course of treatment for a patient, wherein the level of cytosolic BAG-1 protein expression is determined relative to a reference level of BAG-1. Thus, in view of the disclosure of Turner et al. that there was no statistically significant correlation between OS and DDFS for invasive carcinoma patients with overexpression of cytoplasmic BAG-1, and the disclosure of Krajewski et al. that higher levels of BAG-1 nuclear immunostaining correlated with OS among patients with early stage breast cancer, one skilled in the art would not have a reasonable expectation of success in practicing the claimed methods.

Accordingly, Turner et al. as evidenced by Krajewski et al. and Exhibit 2 neither teaches nor suggests the claimed prognostic methods relating to stage I or stage II of breast cancer and the level of cytosolic BAG-1 protein expression. The disclosure of Sano et al. does not cure these defects either, as it contains no teaching or suggestions that would complement the disclosure of Turner et al. evidenced by Krajewski et al. and Exhibit 2 to arrive at the claimed methods. Therefore, Applicant respectfully maintains that the claimed methods are unobvious over Turner et al., alone or in combination with Sano et al. as evidenced by Krajewski et al. and Exhibit 2. Accordingly, Applicant respectfully requests that this rejection be withdrawn.

The rejection of claims 11, 16, 24-27, 32, 34, 44, 50-54, 56, 58-68, 75-79, 81, and 83-109 under 35 U.S.C. § 103(a) as allegedly obvious over Turner et al., *supra*, in view of Sauter et al., Br. J. Cancer 76:494-501 (1997) as evidenced by Krajewski et al., *supra*, and Exhibit 2, *supra*, is respectfully traversed. The Office Action states on page 6 that Applicant's previous arguments are not persuasive for the reasons set forth in the 102(b) rejection.

Applicant respectfully maintains, for the reasons of record, that the claimed methods are unobvious over Turner et al., alone or in combination with Sauter et al. Additionally, as discussed above, Turner et al. as evidenced by Krajewski et al. and Exhibit 2 neither teaches nor suggests the claimed prognostic methods relating to stage I or stage II of breast cancer and the level of cytosolic BAG-1 protein expression. The disclosure of Sauter et al. does not cure these defects either, as it contains no teaching or suggestions that would complement the teaching of

Turner et al. evidenced by Krajewski et al. and Exhibit 2 to arrive at the claimed methods. Therefore, Applicant respectfully maintains that the claimed methods are unobvious over Turner et al., alone or in combination with Sauter et al. as evidenced by Krajewski et al. and Exhibit 2. Accordingly, Applicant respectfully requests that this rejection be withdrawn.

The rejection of claims 11, 16, 24-27, 32, 34, 44, 50-54, 56, 58-68, 75-79, 81, and 83-109 under 35 U.S.C. § 103(a) as allegedly obvious over Turner et al., *supra*, in view of Love (U.S. Patent No. 6,221,622) as evidenced by Krajewski et al., *supra*, and Exhibit 2, *supra*, is respectfully traversed. The Office Action states on page 7 that Applicant's previous arguments are not persuasive for the reasons set forth in the 102(b) rejection.

Applicant respectfully maintains, for the reasons of record, that the claimed methods are unobvious over Turner et al., alone or in combination with Sauter et al. Additionally, as discussed above, Turner et al. as evidenced by Krajewski et al. and Exhibit 2 neither teaches nor suggests the claimed prognostic methods relating to stage I or stage II of breast cancer and the level of cytosolic BAG-1 protein expression. The disclosure of Love does not cure these defects either, as it contains no teaching or suggestions that would complement the teaching of Turner et al. evidenced by Krajewski et al. and Exhibit 2 to arrive at the claimed methods. Therefore, Applicant respectfully maintains that the claimed methods are unobvious over Turner et al., alone or in combination with Love as evidenced by Krajewski et al. and Exhibit 2. Accordingly, Applicant respectfully requests that this rejection be withdrawn.

New Grounds for Rejections

The rejection of claims 11, 16, 22-27, 32, 34, 44, 50-54, 56, 58-61, 67, 68, 73-79, 81, 83-86, 88, 89, 91, 92, 94, 95, 97, 98, and 100-109 under 35 U.S.C. § 103(a) as allegedly obvious over Turner et al., *supra*, in view of Mather et al., Clin. Cancer Res., 4:1857-1856 (1998) and McGuire et al. (US Patent 6,188,964) as evidenced by Krajewski et al., *supra*, and Exhibit 2, *supra*, is respectfully traversed.

As discussed above, Turner et al. as evidenced by Krajewski et al. and Exhibit 2 neither teaches nor suggests the claimed prognostic methods relating to stage I or stage II of breast cancer and the level of cytosolic BAG-1 protein expression. The disclosures of Mather et al.

and/or McGuire et al. do not cure these defects, as they contains no teaching or suggestions that would complement the teaching of Turner et al. evidenced by Krajewski et al. and Exhibit 2 to arrive at the claimed methods. Therefore, Applicant respectfully maintains that the claimed methods are unobvious over Turner et al., alone or in combination with Mather et al. and/or McGuire et al. as evidenced by Krajewski et al. and Exhibit 2. Accordingly, Applicant respectfully requests that this rejection be withdrawn.

The rejection of claims 11, 12, 16, 22-27, 32-34, 44, 50-54, 56, 58-69, 73-81, and 83-109 under 35 U.S.C. § 103(a) as allegedly obvious over Turner et al., *supra*, in view of Mather et al., *supra*, McGuire et al., *supra*, Sano et al., *supra*, and Love, *supra*, is respectfully traversed.

The Office asserts on page 9 that Turner et al. disclose that there is a statistically significant over expression of nuclear and cytoplasmic BAG-1 in cancer patients compared to BBE patients. Applicant respectfully disagrees with this characterization of the disclosure of Turner et al. As discussed above, at best, Turner et al. identify the analysis of OS and DDFS for invasive carcinoma patients with overexpression of cytoplasmic BAG-1 revealed a p value of 0.06. This would most likely indicate to one skilled in the art that there was no statistically significant correlation between OS and DDFS for invasive carcinoma patients with overexpression of cytoplasmic BAG-1. Furthermore, Exhibit A provided herewith corroborates Dr. Reed's statement that the patient survival difference of cytosolic staining of BAG-1 cited in Turner et al. for invasive carcinoma (IC) was not statistically significant ($p = 0.06$) (see Reed Declaration submitted previously as Exhibit 2 with the response filed December 13, 2006). Accordingly, Turner et al. describe a retrospective study that showed no statistically significant correlation between invasive carcinoma patients survival and overexpression of cytoplasmic BAG-1. It is the Applicant, in the present invention, that discovered 1) a statistically significant correlation between 10-year OS and DMFS (distant metastasis-free survival) for patients with overexpression of BAG-1 protein in stages I and II of breast cancer that was 90% and 84%, respectively, as compared with 40% and 40% for those with low BAG-1 levels ($p<0.001$) (see page 35 of the application as filed, lines 5-14, and Figure 1) and 2) patients whose tumors contained high levels of cytosolic BAG-1 protein are more likely to enjoy long-term survival and freedom from tumor recurrence or spread and distant metastases, compared to those with tumors containing low levels of cytosolic BAG-1 (see page 35 of the application as filed, line 31 to page

36, line 3). Thus, one skilled in the art would not have a reasonable expectation of success in practicing the claimed methods.

Applicant further submits, as discussed above, that the invasive carcinoma specimens as taught by Turner et al. would likely be considered by one skilled in the art to include stage I, stage II, stage III, or stage IV breast cancers. This is supported by the evidence of record, specifically Exhibit 1 (a printout from the Breastcancer.org website entitled "Non-Invasive or Invasive Breast Cancer?" (www.breastcancer.org/symptoms/diagnosis/invasive.jsp)), Exhibit 2, *supra*, and Exhibit 3 (Markman, Basic Cancer Medicine pgs. 35-37 (1997)) submitted with the response filed July 23, 2008. Exhibit 1 defines invasive (or infiltrating) breast cancers to be cancers that have started to break through normal breast tissue barriers and invade surrounding area. Exhibit 2 explicitly disclose that breast cancer at stage I, stage II, stage III or stage IV is considered to be invasive. Additionally, according to the tumor-node-metastasis (TNM) staging system described in Exhibit 3, there are thirteen distinguishable TNM stage groupings within stage I to stage IV of breast cancer (see page 36 of Exhibit 3). Thus, Applicant submits that one skilled in the art would understand that invasive carcinoma, as taught by Turner, could include stage I, II, III, or IV breast cancers or any of the thirteen TNM stage groupings. Applicant also submits that, based on the description in Turner et al. and what was well known in the art, one skilled in the art would not know which stage of breast cancer the IC samples of Turner et al. belong and certainly would not know if the samples were stage I or stage II.

Accordingly, Turner et al. neither teaches nor suggests the claimed prognostic methods relating to stage I or stage II of breast cancer and the level of cytosolic BAG-1 protein expression. The disclosures of Mather et al. and/or McGuire et al. and/or Sano et al. and/or Love do not cure these defects, as they contains no teaching or suggestions that would complement the teaching of Turner et al to arrive at the claimed methods. Therefore, Applicant respectfully maintains that the claimed methods are unobvious over Turner et al., alone or in combination with Mather et al. and/or McGuire et al. and/or Sano et al. and/or Love. Accordingly, Applicant respectfully requests that this rejection be withdrawn.

CONCLUSION

In light of the remarks herein, Applicant submits that the claims are now in condition for allowance and respectfully requests a notice to this effect. The Examiner is invited to call the undersigned agent if there are any questions.

To the extent necessary, a petition for an extension of time under 37 C.F.R. 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account 502624 and please credit any excess fees to such deposit account.

Respectfully submitted,

McDERMOTT WILL & EMERY LLP

/Micheal L. Hebert/
Micheal L. Hebert
Registration No. 63,329

11682 El Camino Real, Suite 400
San Diego, CA 92130
Phone: 858.720.3300 MLH:llf
Facsimile: 858.720.7800
Date: May 11, 2010

Please recognize our Customer No. 41552 as our correspondence address.

Exhibit A

BIOSTATISTICAL ANALYSIS

FOURTH EDITION

JERROLD H. ZAR

Department of Biological Sciences
Northern Illinois University

PRENTICE HALL
Upper Saddle River, New Jersey 07458

Library of Congress Cataloging in Publication Data
Zar, Jerrold H.
Biostatistical analysis / Jerrold H. Zar. — 4th ed.
p. cm.
Includes bibliographical references (p.) and index.
ISBN 0-13-081542-X (alk. paper)
1. Biometry. I. Title
QH323.5.Z37 1999
570'.1'5195—dc21 98-34062
CIP

Editorial/production supervision: *Interactive Composition Corporation*
Cover director: *Jayne Conte*
Cover designer: *Bruce Kenselaar*
Manufacturing manager: *Trudy Pisciotti*
Editor: *Teresa Ryu*
Senior editor: *Sheri L. Snavely*
Editorial assistants: *Nancy Bauer* and *Lisa Tarabokjia*

Printed in the United States of America

10 9 8 7 6 5 4 3 2

ISBN 0-13-081542-X

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S. A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

CON

EXAMPLE 6.5 The calculation of the standard error of the mean, $s_{\bar{X}}$. The following are data for systolic blood pressure, in mm of mercury.

121	$n = 12$
125	$\bar{X} = \dfrac{1651 \text{ mm}}{12} = 137.6 \text{ mm}$
128	
134	$SS = 228{,}111 \text{ mm}^2 - \dfrac{(1651 \text{ mm})^2}{12}$
136	
138	$= 960.9167 \text{ mm}^2$
139	$s^2 = \dfrac{960.9167 \text{ mm}^2}{11} = 87.3561 \text{ mm}^2$
141	
144	$s = \sqrt{87.3561 \text{ mm}^2} = 9.35 \text{ mm}$
145	$s_{\bar{X}} = \dfrac{s}{\sqrt{n}} = \dfrac{9.35 \text{ mm}}{\sqrt{12}} = 2.7 \text{ mm}$ *or*
149	
151	
$\sum X = 1651 \text{ mm}$	$s_{\bar{X}} = \sqrt{\dfrac{s^2}{n}} = \sqrt{\dfrac{87.3561 \text{ mm}^2}{12}} = \sqrt{7.2797 \text{ mm}^2} = 2.7 \text{ mm}$
$\sum X^2 = 228{,}111 \text{ mm}^2$	

6.4 Introduction to Statistical Hypothesis Testing

A major goal of statistical analysis is to draw inferences about a population by examining a sample from that population. A very common example of this is the desire to draw conclusions about one or more population means.

We begin by making a concise statement about the population mean, a statement called a *null hypothesis* (abbreviated H_0)* because it expresses the concept of "no difference." For example, a null hypothesis about a population mean (μ) might assert that μ is not different from zero (i.e., μ is equal to zero); and this would be written as

$$H_0\text{: } \mu = 0.$$

Or, we could hypothesize that the population mean is not different from (i.e., is equal to) 3.5 cm, or not different from 10.5 kg, in which case we would write H_0: $\mu = 3.5$ cm or H_0: $\mu = 10.5$ kg, respectively.

If it is concluded that it is likely that a null hypothesis is false, then an *alternate hypothesis* (abbreviated H_A) is assumed to be true. One states a null hypothesis and an alternate hypothesis for each statistical test performed, and all possible outcomes are accounted for by this pair of hypotheses. So, for the examples above:

$$H_0\text{: } \mu = 0, \quad H_A\text{: } \mu \neq 0;$$

$$H_0\text{: } \mu = 3.5 \text{ cm}, \quad H_A\text{: } \mu \neq 3.5 \text{ cm};$$

$$H_0\text{: } \mu = 10.5 \text{ kg}, \quad H_A\text{: } \mu \neq 10.5 \text{ kg}.$$

*E. S. Pearson (1947) credits the introduction of the symbol, "H_0," to J. Neyman and himself and the origin of the term, "null hypothesis," to R. A. Fisher; David (1995) credits a 1935 Fisher paper with introducing the latter term.

It must be emphasized that statistical hypotheses are to be stated *before* data are collected to test them. To propose hypotheses after examination of data can invalidate a statistical test. One may, however, legitimately formulate hypotheses *after* inspecting data if a new set of data is then collected with which to test the hypotheses.

Statistical Testing and Probability. Statistical testing of a null hypothesis about the mean of a population (μ) involves determining the mean of a random sample from that population $\bar{X}$. Then we determine the probability, *if H_0 is true*, of an $\bar{X}$ at least as far from μ as the $\bar{X}$ in the sample. This is accomplished by the considerations of Section 6.3 and is demonstrated in Example 6.6.

Here, a manufacturer has produced a device that is to sound an alarm when the concentration of carbon monoxide (CO) in the air is at 10.00 mg/m^3; and we wish to know whether the device works as intended. Known amounts of carbon monoxide are introduced into a chamber initially containing no CO, and it is recorded at what CO concentration the alarm sounds. This is done eighteen times, with the resultant data (the eighteen values of X_i) shown in Example 6.6. These eighteen data have a mean of $\bar{X} = 10.43$ mg/m^3 and they represent a sample (we presume a random sample) of a very large number of data, namely the very large number of alarm-triggering CO concentrations that would result from repeating this experiment a very large number of times. This large number of X_i's is the statistical population. Although one almost never knows the actual parameters of a sampled population, for this introduction to statistical testing let us suppose that the variance of the population for this example is known to be $\sigma^2 = 1.0434$ (mg/m^3)^2. Thus, for the population of means that could be drawn from this population of measurements, the standard error of the mean is $\sigma_{\bar{X}} = \sqrt{1.0434 \text{ (mg/m}^3)^2/18} = \sqrt{0.0580 \text{ (mg/m}^3)^2} = 0.24$ mg/m^3.

As we wish to know whether the mean of a very large number of repetitions of this experiment equals 10.00 mg/m^3, the appropriate null and alternate hypotheses are H_0: $\mu = 10.00$ mg/m^3 and H_A: $\mu \neq 10.00$ mg/m^3, respectively. And, what we ask is the following:

> If we have a normal population with $\mu = 10.00$ mg/m^3 and $\sigma_{\bar{X}} = 0.24$ (mg/m^3), what is the probability of obtaining a random sample with a mean ($\bar{X}$) at least as far from 10.00 mg/m^3 as 10.43 mg/m^3? Another way to state this would be: What is $P(\bar{X} \geq 10.43 \text{ mg/m}^3 \text{ or } \bar{X} \leq 9.57 \text{ mg/m}^3)$?

Reflecting upon Section 6.3 it can be seen that such probabilities may be ascertained by computation of Z (by Equation 6.16), so Z is referred to as a *test statistic*. The above null hypothesis is tested in Example 6.6, and the two normal-distribution tail regions of interest are shown in Fig. 6.5.

Statistical Errors in Hypothesis Testing. One needs an objective criterion for rejecting or not rejecting the null hypothesis for a statistical test. Theoretically, a very large (or very small) sample mean might be obtained—and a very large absolute value of Z thereby computed—even when H_0 is true; however, the larger the $|Z|$ the smaller the probability that H_0 is true. So we can ask: "How small a probability (or, how

EXAMPLE 6.6 Testing the hypotheses H_0: $\mu = 10.00$ mg/m^3 and H_A: $\mu \neq 10.00$ mg/m^3.

The variable, X, is the carbon monoxide concentration in air, and eighteen measurements are obtained, as follows: 10.25, 10.37, 10.66, 10.47, 10.56, 10.22, 10.44, 10.38, 10.63, 10.40, 10.39, 10.26, 10.32, 10.35, 10.54, 10.33, 10.48, 10.68 mg/m^3.

For these data the sample mean is $\bar{X} = 10.43$ mg/m^3; and for the sake of this example the population standard error of the mean is said to be known to be $\sigma_{\bar{X}} = 0.24$ mg/m^3.

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{10.43 \text{ mg/m}^3 - 10.00 \text{ mg/m}^3}{0.24 \text{ mg/m}^3} = 1.79$$

Using Table B.2:

$$P(\bar{X} \geq 10.43 \text{ mg/m}^3) = P(Z \geq 1.79) = 0.0367$$

and

$$P(Z \leq -1.79) = 0.0367.$$

Therefore,

$$P(\bar{X} \geq 10.43 \text{ mg/m}^3 \text{ or } \bar{X} \leq 9.57 \text{ mg/m}^3) = 0.0367 + 0.0367 = 0.0734.$$

As $0.0734 > 0.05$, do not reject H_0.

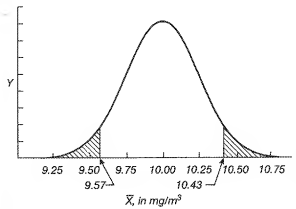

Figure 6.5 The normal distribution of means referred to in the hypothesis testing of Example 6.6a, with a hypothesized population mean, μ, of 10.00 mg/m^3 and a presumed population standard error of the mean, $\sigma_{\bar{X}}$, of 0.24 mg/m^3. It is found that there is a probability of 0.0734 of the mean of a sample of eighteen being in the shaded portions of the two tails.

large a $|Z|$) will be required to reject the null hypothesis?" As explained below, a probability of 5% or less is commonly used as the criterion for rejection of H_0. The probability used as the criterion for rejection is called the *significance level*,* denoted by α (the lowercase Greek letter, alpha). The value of the test statistic (in this case, Z) corresponding to α is termed the *critical value* of the test statistic. In Appendix B.2 it is seen that $P(Z \geq 1.96) = 0.025$; and, inasmuch as the normal distribution is symmetrical, $P(Z \leq -1.96) = 0.025$. Therefore, the critical value for testing the above H_0 at a 5%

*David (1995) credits R. A. Fisher as the first to refer to "level of significance," in 1925, when Fisher also formally recommended use of the 5% level (Cowles and Davis, 1982).

level of significance is $Z = 1.96$. As the test statistic in Example 6.6 (namely, $Z = 1.79$) is not as large as the critical value, the null hypotheses is not rejected.

It is very important to realize that a true null hypothesis occasionally will be rejected, which of course means that we have committed an error in drawing a conclusion about the sampled population. Moreover, this error will be committed with a frequency of α. That is, if H_0 is in fact a true statement about a statistical population, it will be concluded (erroneously) to be false 5% of the time. The rejection of a null hypothesis when it is in fact true is what is known as a *Type I error* ("Type 1 error," also called an alpha error or an "error of the first kind"). On the other hand, if H_0 is in fact false, a statistical test will sometimes not detect this fact, and we shall thus reach an erroneous conclusion by not rejecting H_0. The probability of committing this error, of not rejecting the null hypothesis when it is in fact false, is represented by β (the lowercase Greek letter, beta). This error is referred to as a *Type II error* ("Type 2 error," also called a beta error, or an "error of the second kind"). The *power* of a statistical test is defined as $1 - \beta$; i.e., power is the probability of rejecting the null hypothesis when it is in fact false and should be rejected.*

Whereas the probability of committing a Type I error is α, the specified significance level, the probability of committing a Type II error is β, a value that generally we neither specify nor know. What we do know is that for a given sample size, n, the value of α is related inversely to the value of β. That is, lower probabilities of committing a Type I error are associated with higher probabilities of committing a Type II error. Both types of error may be reduced simultaneously by increasing n. Thus, for a given α, larger samples will result in statistical testing with greater power $(1 - \beta)$.

Table 6.1 summarizes these two types of statistical errors. Since, for a given n, one cannot minimize both of them, it is appropriate to ask what the acceptable combination of the two might be. By experience, and hence by convention, an α of 0.05 is usually considered to be a "small enough" chance of committing a Type I error, while not being so small as to result in "too large a chance" of a Type II error. But there is nothing sacrosanct about the 0.05 level. Although it is the most widely used significance level, researchers may decide for themselves whether it is more important to minimize one type of error or the other. In some situations, for example, a 5% chance of an incorrect rejection of H_0 may be felt to be unacceptably high, so the 1% level of significance is sometimes employed. It is necessary, of course, to state the significance level used when reporting the results of a statistical test. Indeed, rather than simply stating whether the null hypothesis is rejected, it is good practice to state also the test statistic itself and the best estimate of its exact probability. (In Example 6.6, it is asserted that $Z = 1.79$ and $P = 0.0734$, in addition to expressing the conclusion that H_0 is not rejected.) In this way, readers of the research results may draw their own conclusions, even if their choice of significance level is different from the author's. Bear in mind, however, that

*The distinction between these two fundamental kinds of statistical errors, and the concept of power, date back to the pioneering work, in England, of Jerzy Neyman (1894–1981; Russian-born, of Polish roots, emigrating as an adult to Poland and then to England, and spending the last half of his life in the United States) and the English statistician, Egon S. Pearson (1895–1980) (Lehmann and Reid, 1982; Neyman and Pearson, 1928a; Pearson, 1947). Their naming of the two types of errors, and of power, dates from 1933 (David, 1995).

TABLE 6.1 The Two Types of Errors in Hypothesis Testing

	If H_0 is true	If H_0 is false
If H_0 is rejected:	Type I error	No error
If H_0 is not rejected:	No error	Type II error

the choice of α is to be made before even seeing the data. Otherwise there is a great risk of having the choice influenced by examination of the data, introducing bias instead of objectivity into the proceedings. The best practice generally is to decide on the null hypothesis, alternate hypothesis, and significance level before commencing with data collection. It is conventional to refer to rejection of H_0 at the 5% significance level as denoting a "significant" difference (in the present example, a significant difference between $\bar{X}$ and 10.00 mg/m^3) and rejection at the 1% level as indicating a "highly significant difference."*

As the significance level selected is somewhat arbitrary, if test results are very near that level (e.g., between 0.04 and 0.06 if $\alpha = 0.05$ is used), then it may be wiser to repeat the analysis with additional data than to declare emphatically that the null hypothesis is or is not a reasonable statement about the sampled population.

More on Statistical Power. The power of a statistical testing procedure was defined above as the probability that the test will correctly reject the null hypothesis when it is false. For many hypothesis tests it is possible to compute the power under specified conditions. For the situation in Example 6.6, for instance, it was hypothesized that the population mean, μ, is 10.00 mg/m^3; and we could ask how powerful the test would be in rejecting H_0 if in reality H_0 is 10.50 mg/m^3. The normal curve in Fig. 6.6a has a mean of 10.50 mg/m^3 and the same standard error as the curve in Fig. 6.5. If, in the sampled population, $\mu = 10.50$ mg/m^3, then the power of the testing procedure of Example is the proportion of the area under the curve of Fig. 6.6a that lies below 9.57 mg/m^3 and above 10.43 mg/m^3. This is computed in Example 6.7a to be 0.61, which means that there was only a 61% chance of rejecting H_0 even though it is false. Fig. 6.6b presents the normal curve of Fig. 6.6a (for which the mean of 10.50 mg/m^3 exists in the population that was sampled) together with the normal curve of Fig. 6.5 (for which the mean of 10.00 mg/m^3 is what was hypothesized), so the relationship between the two may be more clearly observed.

If the population mean, μ, is really larger than 10.50 mg/m^3, then the curve would be ever farther to the right than in Fig. 6.6a and the power would be greater than 0.61, as shown in Example 6.7b. This demonstrates the general principle that the farther a population characteristic is from that stated in the null hypothesis, the greater will be the power to reject the null hypothesis.

*Many authors attach an asterisk (*) to a test statistic if it is associated with a probability ≤ 0.05 and two asterisks (**) if the probability is ≤ 0.01.

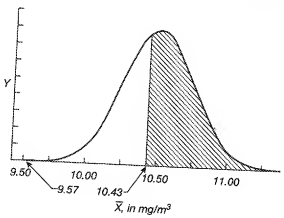

Figure 6.6a A normal distribution of means with a mean of 10.50 mg/m³ and the same standard error as in Fig. 6.5. The shaded area (61% of the area under the curve) is the power of the test performed in Example 6.6b.

One-Tailed Testing. As in Example 6.6, we have thus far considered the testing of null hypotheses declaring no difference between a population parameter and a hypothesized value, where the alternate hypothesis embodies difference in either of two directions from that value. There are instances, however, where the researcher's interest is in whether there is in the population a difference *in a specified direction* between a population parameter and a hypothesized value. Indeed, the data in Example 6.6 might have been collected with the express purpose of asking whether the mean CO concentration that sets off the alarm is *greater than* (rather than simply different from) 10.00 mg/m³. As the null hypothesis is to contain the concept of "no difference," the appropriate H_0 and H_A may be stated as

$$H_0\text{: } \mu \leq 10.00 \text{ mg/m}^3;$$

$$H_A\text{: } \mu > 10.00 \text{ mg/m}^3.$$

These hypotheses would be preferred to those in Example 6.6 if, for example, the measurement device is to be employed in environmental-safety monitoring and our concern is whether the carbon monoxide concentration *exceeds* a specified level (namely, 10.00 mg/m³), which is expressed in H_A. The above null hypothesis is rejected if the calculated Z is in the extreme α (e.g., 5%) of the *right-hand tail* (i.e., if Z is positive and $P(Z) \leq \alpha$). For the data in Example 6.6, this probability is 0.0367; and, as this is < 0.05, H_0 is rejected.

If, instead, our interest were specifically whether the mean of the sampled population is *less than* a hypothesized value, then the relevant statistical hypotheses would be

$$H_0\text{: } \mu \geq 10.00 \text{ mg/m}^3;$$

$$H_A\text{: } \mu < 10.00 \text{ mg/m}^3;$$

and H_0 would be rejected only if the calculated Z is in the extreme α (e.g., 5%) of the *left-hand tail* (i.e., if Z is *negative* and $P(Z) \leq \alpha$).

Statistical testing that examines differences in only one of two possible directions is called *one-tailed testing*. It may be seen in Appendix B.2 that, if one employs the

EXAMPLE 6.7 Calculating the power of the hypothesis test of Example 6.6.

H_0: $\mu = 10.00$ mg/m^3; H_A: $\mu \neq 10.00$ mg/m^3; $\sigma_{\bar{X}} = 0.24$ mg/m^3.

(a) What is the power if the population mean is actually 10.50 mg/m^3?

$$\text{Power} = P(\bar{X} \leq 9.57 \text{ mg/m}^3 \text{ or } \bar{X} \geq 10.43 \text{ mg/m}^3)$$

For $P(\bar{X} \leq 9.57$ mg/m$^3)$:

$$Z = \frac{9.57 \text{ mg/m}^3 - 10.50 \text{ mg/m}^3}{0.24 \text{ mg/m}^3} = -3.88$$

Consulting Table B.2:

$$P(Z \leq -3.88) = P(Z \geq 3.88) = 0.0001$$

For $P(\bar{X} \geq 10.43$ mg/m$^3)$:

$$Z = \frac{10.43 \text{ mg/m}^3 - 10.50 \text{ mg/m}^3}{0.24 \text{ mg/m}^3} = -0.29$$

$$P(Z \geq -0.29) = 1 - P(Z \leq -0.29) = 1 - P(Z \geq 0.29) = 1 - 0.3859 = 0.6141$$

So, power $= 0.0001 + 0.6141 = 0.6142$.

(b) What is the power if the population mean is actually 10.75 mg/m^3?

$$\text{Power} = P(\bar{X} \leq 9.57 \text{ mg/m}^3 \text{ or } \bar{X} \geq 10.43 \text{ mg/m}^3)$$

For $P(\bar{X} \leq 9.57$ mg/m$^3)$:

$$Z = \frac{9.57 \text{ mg/m}^3 - 10.75 \text{ mg/m}^3}{0.24 \text{ mg/m}^3} = -4.92$$

$$P(Z \leq -4.92) = P(Z \geq 4.92) = 0.0000$$

For $P(\bar{X} \geq 10.43$ mg/m$^3)$:

$$Z = \frac{10.43 \text{ mg/m}^3 - 10.75 \text{ mg/m}^3}{0.24 \text{ mg/m}^3} = -1.33$$

$$P(Z \geq -1.33) = 1 - P(Z \leq -1.33) = 1 - P(Z \geq 1.33) = 1 - 0.0918 = 0.9082$$

So, power $= 0.0000 + 0.9082 = 0.9082$.

5% level of significance, the one-tailed critical value of Z is 1.645. The one-tailed critical value (let's call it $Z_{\alpha(1)}$) is always less than the two-tailed critical value ($Z_{\alpha(2)}$); for example, for the 5% significance level $Z_{\alpha(1)} = 1.645$ and $Z_{\alpha(2)} = 1.960$. Thus, a researcher may be tempted to use one-tailed testing if he or she wishes to declare significance and the calculated Z lies between $Z_{\alpha(1)}$ and $Z_{\alpha(2)}$ (as in Example 6.6). But, this would not be a valid use of this testing procedure, for recall that *statistical hypotheses are to be declared before examining the data*, and they should reflect the question of interest about the population. Another example of one-tailed testing of a mean is found in Exercise 6.6, parts b and c.

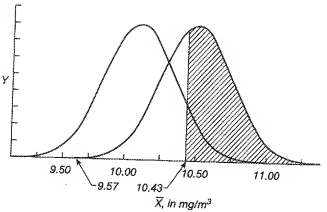

Figure 6.6b The normal distributions of Figures 6.5 and 6.6a.

6.5 Assessing Departures from Normality

It is sometimes desired to test the hypothesis that a sample came from a population whose members follow a normal distribution. Example 6.1 and Fig. 6.7 present a frequency distribution of sample data, and we may wish to ask if the data came from a population that is normally distributed.

"Outliers." Occasionally a set of data will have one or more observations that are so extreme in value, relative to the other data in the sample, that we suspect they should not be a part of the sample. Such data are often called "outliers." Sometimes it is clear that an outlier is the result of incorrect recording of data. For example, if Example 6.1 contained a student height of 680 inches, or 7.2 inches, or 160 inches, we would immediately suspect an error. We might surmise that there was a decimal-point error in recording the first two (which might have actually been heights of 68 inches and 72 inches, respectively) and that the third height was measured in centimeters and mistakenly recorded as if it were inches. In other instances it is known that a measurement was faulty (e.g., a measuring instrument malfunctioned or a technician contaminated the item to be measured).

It is not appropriate to discard data simply because they appear (to someone) to be unreasonably extreme. However, there might be a very obvious reason such as the above for correcting or discarding a datum. If there are not such conspicuous grounds, then an outlier might be objectively rejected as erroneous by statistical methods that are referred to in the later discussions on comparing variables among populations (the end of Section 10.1) or analyzing the relationship between variables (the end of Section 16.2).

Goodness of Fit Testing of Normality. Goodness of fit procedures for testing the null hypothesis that a sample came from a normal population are based on the methods to be discussed in Chapter 21. Methods such as chi-square, or Kolmogorov-Smironov goodness of fit procedures, with appropriate modifications, can be used to test the hypothesis of normality in the population distribution (e.g., Zar, 1984: 88–93). These